# 臺灣大學應用數學科學研究所 107 學年度碩士班甄試試題

## 科目：機率統計

2017.10.20

1. (15%) $X$ and $Y$ are a pair of random variables. Suppose the joint density function of $(X, Y)$ is $f(x, y) = c$ for $(x-1)^2 + (y-3)^2 = 2$, and is otherwise 0. Determine $c$, $E(X)$, and $Cov(X, Y)$. (Note: Five points each.)

2. (20%) (The $t_1$ distribution)

   **(a)** (10%) Let $T \sim t_1$ (the $t$ distribution with 1 degree of freedom). Explain why $T$ has the same distribution as $X/|Y|$. Here $X$ and $Y$ are independent and normally distributed random variable with mean 0 and variance 1.

   **(b)** (10%) For $T \sim t_1$, show that $E(|T|) = \infty$ and $E(T^2) = \infty$. If $T_1, \cdots, T_n$ are independent and identically distributed random variables coming from distribution $t_1$, explain why the Law of Large Numbers and the Central Limit Theorem do not apply to the sample mean $(T_1 + \cdots + T_n)/n$.

   You may use this following result without proof in part (b). The density function of Cauchy random variable is
   $$f(x) = \frac{1}{\pi}\frac{1}{x^2+1}.$$

3. (20%) (Comparing binomial proportions.) The internet company WHO would like to understand whether visitors to a website are more likely to click on an advertisement at the top of the page than one on the side of the page. They conduct an *AB test* in which they show $n$ visitors (group $A$) a version of the website with the advertisement at the top, and $m$ visitors (group $B$) a version of the website with the (same) advertisement at the side. They record how many visitors in each group clicked on the advertisement.

   **(a)** (5%) Formulate this problem as a *hypothesis test problem.* (You may assume that visitors in group $A$ independently click on the ad with probability $p_A$ and visitors in group $B$ independently click on the ad with probability $p_B$, where both $p_A$ and $p_B$ are unknown probabilities in $(0, 1)$.) What are the null and alternative hypotheses? Are they *simple* or *composite*?

   **(b)** (5%) Let $\hat{p}_A$ be the fraction of visitors in group $A$ who clicked on the ad, and similarly for $\hat{p}_B$. A reasonable intuition is to reject $H_0$ when $\hat{p}_A - \hat{p}_B$ is large. What is the variance of $\hat{p}_A - \hat{p}_B$? Is this the same for all data distributions in $H_0$?

   **(c)** (10%) Describe a way to estimate the variance of $\hat{p}_A - \hat{p}_B$ using the available data, assuming the null hypothesis $H_0$ is true-call this estimate $\hat{V}$. Explain heuristically why, when $n$ and $m$ are both large, the test statistic
   $$T = (\hat{p}_A - \hat{p}_B)/\sqrt{\hat{V}}$$
   is approximately distributed as $N(0, 1)$ under any data distribution in $H_0$. (You may assume that when $n$ and $m$ are both large, the ratio of $\hat{V}$ to the true variance of $\hat{p}_A - \hat{p}_B$ that you derived in part (b) is very close to 1 with high probability.) Explain how to use this observation to perform an approximate level-$\alpha$ test of $H_0$ versus $H_1$.

4. (25%) (Monte Carlo integration) For a given function $f : [a, b] \to R$, suppose we wish to numerically evaluate

$$I(f) = \int_a^b f(x)dx.$$

One method is the following: Let $g$ be a probability density function of a continuous random variable taking values in $[a, b]$, and generate independent random draws $X_1, \cdots, X_n$ from $g$. Then estimate $I(f)$ by

$$\hat{I}_n(f) = \frac{1}{n} \sum_{i=1}^{n} \frac{f(X_i)}{g(X_i)}.$$

**(a)** (5%) Show that $E[\hat{I}_n(f)] = I(f)$.

**(b)** (5%) Assuming that $Var(f(X_i)/g(X_i)) < \infty$, show that $\hat{I}_n(f) \to I(f)$ in probability as $n$ goes to the infinity.

**(c)** (10%) Derive a formula for $Var(\hat{I}_n(f))$. Determine $c_n \in R$ to ensure $c_n(\hat{I}_n(f) - I(f)) \to N(0, 1)$ in distribution as $n$ goes to the infinity.

**(d)** (5%) Consider concretely the problem of evaluating

$$I(f) = \int_0^1 \cos(2\pi x)dx.$$

Let $g$ be the probability density function of the uniform distribution on $[0, 1]$, and consider the above estimate $\hat{I}_n(f)$ using 1000 independently and identically distributed samples from $g$. Using the result from part (c), compute approximately the probability $P(|\hat{f}_n - I(f)| > 0.05)$.

5. (20%) Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with density $f(\cdot)$. Suppose that $P(X_i \geq 0) = 1$ and that $\lambda = \lim_{x \to 0+} f(x)/x = 1$. Set $X_{(1)}$ to be $\min\{X_1, \cdots, X_n\}$ and $Y_n = nX_{(1)}$. Determine the asymptotic distribution of $Y_n$ as $n$ goes to the infinity.